DOCUMENT RESUME

ED 348 381 TM 018 755

AUTHOR	Gordon, Edmund W.
TITLE	Implications of Diversity in Human Characteristics for Authentic Assessment.
INSTITUTION	Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO	CSE-TR-341
PUB DATE	Apr 92
CONTRACT	R117G10027
NOTE	11p.
PUB TYPE	Viewpoints (Opinion/Position Papers, Essays, etc.) (120)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Academic Achievement; *Academic Standards; *Educational Assessment; Elementary Secondary Education; Hypermedia; *Individual Differences; Minority Groups; Outcomes of Education; Racial Discrimination; Sampling; Sex Bias; *Student Evaluation; *Test Bias; Test Items
IDENTIFIERS	*Authentic Assessment; *National Standards; Performance Based Evaluation

ABSTRACT

The question of the appropriateness of a set of national standards for educational achievement in the United States is extremely complex. One source of the complexities is the diversity of human characteristics, and the distortions that come from racism, sexism, and other forms of chauvinism. The current practice in testing seems to treat validity through an assumptive bias. Commonalty or heterogeneity is simply assumed, without determining how these assumptions are influenced by racism, sexism, classism, or, in the case of language, nationalism. A concern for population diversity and pluralistic outcomes is a necessity in the adequate assessment of achievement. The challenge to authentic assessment is to understand and agree that the concern with diversity, pluralism, and equity rests on commitment to universal standards of competence. Assessment probes should provide: (1) diversity in task content, contexts, demands, and referents; (2) flexibility in timing; (3) multiplicity of perspectives; (4) critical sampling; (5) hypertext, requiring supplying absent elements for embedded knowledge; (6) choice of self-selected and teacher-selected options; (7) opportunities to identify examples of canonical knowledge and technique; (8) individual and cooperative performance opportunities; and (9) self-designated tasks from examinee generated inventories of knowledge, skill, and understanding. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED348381

CRESST

National Center for Research on Evaluation, Standards, and Student Testing

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Implications of Diversity in Human Characteristics for Authentic Assessment

CSE Technical Report 341

Edmund W. Gordon,
John M. Musser Professor Emeritus

Yale University

TM018755

BEST COPY AVAILABLE

▸ UCLA Center for the Study of Evaluation

in collaboration with:

▸ University of Colorado

▸ NORC, University of Chicago

▸ LRDC, University of Pittsburgh

▸ The RAND Corporation

Implications of Diversity in Human Characteristics for Authentic Assessment

CSE Technical Report 341

Edmund W. Gordon,
John M. Musser Professor Emeritus

Yale University

April 1992

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
UCLA Graduate School of Education
University of California, Los Angeles
Los Angeles, CA 90024-1522
(310) 206-1532

The work reported herein was partially supported under the Educational Research and Development Center Program cooperative agreement R117G10027 and CFDA catalog number 84.117G as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the position or policies of the Office of Educational Research and Improvement or the U.S. Department of Education.

Implications of Diversity in Human Characteristics for Authentic Assessment

Edmund W. Gordon, John M. Musser Professor Emeritus

Yale University

The current debate concerning the appropriateness of a set of national standards for educational achievement in the United States has been coupled with a renewed debate concerning the utility of standardized testing (a) in the monitoring of educational progress and (b) as a basis for credentialing for a wide variety of purposes. Unfortunately, for those of us who prefer to deal with simple problems, this one is complex beyond measure.

One source for these complexities is the ubiquitous distortions which flow from the fact of racism, sexism, and other forms of chauvinism in our society. These distortions have been traditionally thought to be unrelated to the processes of education and educational assessment. This is because the tendency has been to focus on the impact of racism and sexism on the persons who are the targeted victims of such communicentric bias and not on the social processes and institutions which reflect those biases. But all of us and all segments of our society are victims or possible victims. The distortions and otherwise negative fallouts have an impact on practically all that we seek to do. Nowhere is this more obvious than in our efforts to educate diverse populations and to assess the educational needs and outcomes in people whose life conditions, experiences and values differ from those which have achieved hegemony in the society.

It is to the credit of many of the recent efforts at reform in the psychometric community that several of us have agreed to try to engage seriously the possible implications of diversity in human characteristics for a more useful and hopefully equitable assessment technology. We seem to have agreed to try to make assessment procedures more authentic with respect to what we know about learning and human competence, as well as with respect to the various populations whose members will be assessed. The concern with authenticity has focused on the development of assessment probes which

require performance, that is, that the respondent do things to demonstrate competence and understanding as in solving problems or explaining relationships. However, this shift from more static to performance measures may not be sufficiently responsive to the diversity in human populations. Our concern for authenticity also requires that we recognize that these various populations live their lives in multiple contexts, and that authenticity may vary not only with populations but also with contexts. Thus, in modern societies authenticity requires that competence be measured by multiple criteria met within the same person functioning in multiple contexts.

In an earlier period a concern for authentic assessment would probably have referred to a concern for validity, reliability and attention to ensuring that standardized procedures were adhered to. But today we are likely to be concerned with more complex psychometric problems long known to confound assessment processes, among which are the problems of test bias, which have been dealt with to the satisfaction of many psychometricians but which continue to frustrate some educators and advocates for civil rights. Others of us are debating questions as to whether to use standardized tests at all. The argument advanced is that traditional standardized items tend to misrepresent the changing nature of knowledge and the processes by which it is acquired and utilized. Some of the most negative critics argue that these procedures and tests penalize not only our weakest students, but many of our most creatively intelligent members. Some of us are ready to concede the importance of some measures of what persons know and know how to do, but insist that it must be possible to develop assessment procedures which are a more appropriate reflection of the ways in which people think, learn, and work, and that are less dependent on recall and regurgitation.

There are those of us who are sympathetic to standards and assessment, but insist that it is immoral to begin by measuring outcomes before we have seriously engaged the equitable and sufficient distribution of inputs, that is, opportunities and resources essential to the development of intellect and competence. So we confront the questions of testing in the face of psychometric, pedagogical, political, economic, psychological, cultural, and philosophical problems, and there appear to be few who are prepared to engage such complex problems from these several perspectives. Not only is there the

tendency to approach these issues from single disciplines but also from personal and sometimes hegemonic identities.

In general, my reference to hegemonic identity relates to the "ubiquitously" distorting effect of communicentric bias—the tendency to see the world from the perspective of my narrow group membership and interest, and to generalize from that truncated perspective to other communities and their peoples. This communicentric bias is reflected in our approaches to class, culture, ethnicity, gender, and language. Spokespersons for the interests of these groups have begun to remind us that such sources of identity and socialization have important influences on the development of the character of adaptive functions and learning. They argue that such variables influence what we learn, the opportunity to learn, how persons store and retrieve information, the motivation to produce and utilize mental products, and more. If they are correct, and I think that they are, these variables have to influence our conceptions of competence with respect to knowledge, intellect and technique. I assume that none of us would argue that these factors do not influence understanding and judgement, my candidates for cognitive products in their highest forms.

Many years ago when the late Bob Thorndike and the marvelous Anne Annistasi were trying to teach me psychometric theory, they insisted that we take questions of validity seriously. In one of the last talks I heard Anne give, she was complaining that psychometric theorists and technologists seem to have given up on the validity question. The current practice seems to treat validity through assumptions: I say through assumptive bias. We simply assume commonality or heterogeneity, but how are these assumptions influenced by racism, sexism, classism, or in the case of language, nationalism?

In efforts at better understanding the influence of these assumptions and our communicentric biases on their development, my colleagues and I have begun the examination of issues related to human diversity and pluralism in society for their implications for the achievement of a higher degree of equity and justice in educational assessment.

We begin with the conviction that it is desirable that attention be given to questions of equity early in the development of an assessment process rather

than as an add-on near the end of such work. Since this issue is more complicated than is often reflected in the public debates, it may be useful to identify some of the possible ways in which a concern for population diversity and pluralistic outcomes impact upon development and learning. It is becoming more and more obvious that these sources of variance influence:

1. the motivation to engage academic learning and to master its content;
2. opportunities to learn and be reinforced by academic competence and literacy;
3. the conditions in and under which knowledge is learned and attitudes are developed toward the disciplines; and
4. the nature of the processes by which academic attitudes, knowledges, competencies, and skills are assessed.

These and other adaptive behaviors are certainly influenced by such social divisions as race and gender. But an exclusive focus on racism or sexism may be less useful for our purposes. Instead, a focus on the implications of diversity and pluralism might better enable us to address the relevant concerns. I refer to class, cultural, ethnic, gender, and language diversity, all of which are possible influences on the manner in which knowledge is acquired and the manner in which academic attitudes and knowledge are produced in assessment demands. It is not clear how much attention we should give to other aspects of diversity such as cognitive style, motivation, and temperament. What is clear is that if we are to adequately assess achievement and to use that assessment information to improve education, we will need to find ways in which to appropriately accommodate relevant sources of diversity in our revised examinations.

A related issue concerns the question of pluralism, that is, the requirement that our students are expected to meet different standards in the multiple contexts in which they live their lives. Obviously, purposes, perspectives, and goals influence what is learned as well as what one is willing to produce. Thus, the assessment problems relate to the appropriateness of the examination probes to the purposes, goals, and standards of the person being examined. In addition, there needs to be concern for the appropriateness of the examination to a context that is pluralistic, as well as for the extent to which the standard context can be made to

accommodate these pluralistic ends without distorting the purposes which are to be served by the examination. Ideally, we want our students to be able to function in multiple contexts and to meet multiple standards. The fact of pluralism in our society makes that necessary; however, current approaches to assessment do not address this problem explicitly.

It is not by accident that existing approaches to standardized assessment are insufficiently sensitive to diversity and pluralism. Dominant standards by which academic competence is judged are calibrated in large measure against either (a) what most persons at a specific level can do, or (b) what we agree is necessary in order for one to take on the next level of work. The fact that some persons have greater difficulty than others or seem unable to achieve that level is thought to be a problem of person characteristics and not a problem with the appropriateness of the measurement or discipline.

In our efforts at being responsive to diverse human characteristics and plural social standards, there may be limits to what can be done in the design and development of assessment procedures. We may be able to make the assessment process more instructive and supportive of instruction. We may find varied contexts in and vehicles through which students can demonstrate their competencies. Our items could be made more process sensitive and give less emphasis to product. But in the final analysis, the assessment procedure is most likely to reveal the effectiveness of the teaching and learning which has occurred. Thus, the facts of diversity and pluralism may have more serious implications for teaching and learning than for assessment. However, this differential in favor of teaching and learning does not eliminate the assessment community's responsibility to be responsive to the facts, problems, and challenges of diversity and pluralism.

This is the challenge to authentic assessment, however, it is essential that we understand and agree that this concern with diversity, pluralism and equity rests upon a commitment to *universal standards of competence,* that is, the same standards for all populations, even though we may be able to agree upon *differential indicators* of change or progress toward those standards. Standards or criteria for competence or mastery cannot be based upon different entry or exit characteristics of learners. Population specific norms may be useful in planning pedagogical intervention, but are irrelevant to certification. Yet, if we are to measure progress, our instruments must be sensitive to

changes within specific populations. The task then is to find assessment probes (test items) which measure the same criterion from contexts and perspectives which reflect the life space and values of the learner. Our indicators must be valid with respect to the criterion used and must be capable of eliciting culturally indigenous behaviors which may reflect incremental movement toward the chosen criterion. To do this will require that we find ways to provide students with learning and testing opportunities which are appropriate to the standard, equivalent to the standard, and sufficient to evoke a relevant response. These may be approached through attention to the engagement potential and interest power of our probes, through the relevance of reference points, and the capacity of items and tasks to be mapped on the learner's existing schema, style, or response repertoires.

This kind of fluidity or flexibility in our probes will require that we come to some agreement concerning the core *knowledges, skills* and *understanding* which are fundamental to developed intellect and then permit some *choice* to the examinee and examiner with respect to how and in which knowledge, skill, and understanding subdomains the examinee demonstrates her or his competence.

Thus *options* and *choices* become a critical feature in any assessment system created to be responsive to equity, just as *processual description* and *diagnosis* become central purposes. There follow a few examples of what our assessment probes should provide:

1. Diversity in task content, contexts, demands and referents;
2. Flexibility in timing entry points, time span of performance, etc.;
3. Multiplicity in perspectives with required comparison and justification;
4. Critical sampling from canonical and noncanonical information and technique;
5. Hypertext: imbedded substantive or procedural knowledge with the requirement that the absent element be provided;
6. Choice involving self-selected and teacher selected options;
7. Opportunity to identify in the indigenous experience examples of canonical knowledge and technique;

8. Individual and cooperative performance opportunities; and

9. Self-designated tasks from examinee generated inventories of knowledge, skill, and understanding: What do I know and how do I choose to demonstrate it?